What is claimed is:

1. An apparatus for converting an input voice signal into an output voice signal according to a target voice signal, the apparatus comprising:

an input device that provides the input voice signal composed of an original sinusoidal component and an original residual component other than the original sinusoidal component;

an extracting device that extracts original attribute data from at least the sinusoidal component of the input voice signal, the original attribute data being characteristic of the input voice signal;

a synthesizing device that synthesizes new attribute data based on both of the original attribute data derived from the input voice signal and target attribute data being characteristic of the target voice signal composed of a target sinusoidal component and a target residual component other than the sinusoidal component, the target attribute data being derived from at least the target sinusoidal component; and

an output device that operates based on the new attribute data and either of the original residual component and the target residual component for producing the output voice signal.

2. The apparatus according to claim 1, wherein the extracting device extracts the original attribute data containing at least one of amplitude data representing an amplitude of the input voice signal, pitch data representing a pitch of the input voice signal, and spectral shape data representing a spectral shape of the input voice signal.

3. The apparatus according to claim 2, wherein the extracting device extracts the original attribute data containing the amplitude data in the form of static amplitude data representing a basic variation of the amplitude and vibrato-like amplitude data representing a minute variation of the amplitude, superposed on the basic variation of the amplitude.

4. The apparatus according to claim 2, wherein the extracting device extracts the original attribute data containing the pitch data in the form of static pitch data representing a basic variation of the pitch and vibrato-like pitch data representing a minute variation of the pitch, superposed on the basic variation of the pitch.

5. The apparatus according to claim 1, wherein the synthesizing device operates based on both of the original attribute data composed of a set of original attribute data elements and the target attribute data composed of another set of target attribute data elements in correspondence with

one another to define each corresponding pair of the original attribute data element and the target attribute data element, such that the synthesizing device selects one of the original attribute data element and the target attribute data element from each corresponding pair for synthesizing the new attribute data composed of a set of new attribute data elements each selected from each corresponding pair.

6. The apparatus according to claim 1, wherein the synthesizing device operates based on both of the original attribute data composed of a set of original attribute data elements and the target attribute data composed of another set of target attribute data elements in correspondence with one another to define each corresponding pair of the original attribute data element and the target attribute data element, such that the synthesizing device interpolates with one another the original attribute data element and the target attribute data element of each corresponding pair for synthesizing the new attribute data composed of a set of new attribute data elements each interpolated from each corresponding pair.

7. The apparatus according to claim 1, further comprising a peripheral device that provides the target attribute data containing pitch data representing a pitch of the target voice signal at a standard key, and a key control device that operates when a user key different than the standard key is

designated to the input voice signal for adjusting the pitch data according to a difference between the standard key and the user key.

8. The apparatus according to claim 1, further comprising a peripheral device that provides the target attribute data divided into a sequence of frames arranged at a standard tempo of the target voice signal, and a tempo control device that operates when a user tempo different than the standard tempo is designated to the input voice signal for adjusting the sequence of the frames of the target attribute data according to a difference between the standard tempo and the user tempo, thereby enabling the synthesizing device to synthesize the new attribute data based on both of the original attribute data and the target attribute data synchronously with each other at the user tempo designated to the input voice signal.

9. The apparatus according to claim 8, wherein the tempo control device adjusts the sequence of the frames of the target attribute data according to the difference between the standard tempo and the user tempo, such that an additional frame of the target attribute data is filled into the sequence of the frames of the target attribute data by interpolation of the target attribute data so as to match with a sequence of frames of the original attribute data provided from the extracting device.

10. The apparatus according to claim 1, further comprising a synchronizing device that compares the target attribute data provided in the form of a first sequence of frames with the original attribute data provided in the form of a second sequence of frames so as to detect a false frame that is present in the second sequence but is absent from the first sequence, and that selects a dummy frame occurring around the false frame in the first sequence so as to compensate for the false frame, thereby synchronizing the first sequence containing the dummy frame to the second sequence containing the false frame.

11. The apparatus according to claim 1, wherein the synthesizing device modifies the new attribute data so that the output device produces the output voice signal based on the modified new attribute data.

12. The apparatus according to claim 1, wherein the synthesizing device synthesizes additional attribute data in addition to the new attribute data so that the output device concurrently produces the output voice signal based on the new attribute data and an additional voice signal based on the additional attribute data in a different pitch than that of the output voice signal.

13. An apparatus for converting an input voice signal into an output voice signal according to a target voice signal, the apparatus comprising:

an input device that provides the input voice signal composed of original sinusoidal components and original residual components other than the original sinusoidal components;

a separating device that separates the original sinusoidal components and the original residual components from each other;

a first modifying device that modifies the original sinusoidal components based on target sinusoidal components contained in the target voice signal so as to form new sinusoidal components having a first pitch;

a second modifying device that modifies the original residual components based on target residual components contained in the target voice signal other than the target sinusoidal components so as to form new residual components having a second pitch;

a shaping device that shapes the new residual components by removing therefrom a fundamental tone corresponding to the second pitch and overtones of the fundamental tone; and

an output device that combines the new sinusoidal components and the shaped new residual components with each other for producing the output voice signal having the first pitch.

14 The apparatus according to claim 13, wherein the shaping device removes the fundamental tone corresponding to the second pitch which is identical to one of a pitch of the original sinusoidal components, a pitch of the target sinusoidal components, and a pitch of the new sinusoidal components.

15 The apparatus according to claim 13, wherein the shaping device comprises a comb filter having a series of peaks of attenuating frequencies corresponding to a series of the fundamental tone and the overtones for filtering the new residual components along a frequency axis.

16 The apparatus according to claim 13, wherein the shaping device comprises a comb filter having a delay loop creating a time delay equivalent to an inverse of the second pitch for filtering the residual components along a time axis so as to remove the fundamental tone and the overtones.

17. An apparatus for converting an input voice signal into an output voice signal according to a target voice signal, the apparatus comprising:

an input device that provides the input voice signal composed of original sinusoidal components and original residual components other than the original sinusoidal components;

a separating device that separates the original sinusoidal components and the original residual components from each other;

a first modifying device that modifies the original sinusoidal components based on target sinusoidal components contained in the target voice signal so as to form new sinusoidal components;

a second modifying device that modifies the original residual components based on target residual components contained in the target voice signal other than the target sinusoidal components so as to form new residual components;

a shaping device that shapes the new residual components by introducing thereinto a fundamental tone and overtones of the fundamental tone corresponding to a desired pitch; and

an output device that combines the new sinusoidal components and the shaped new residual components with each other for producing the output voice signal.

18 The apparatus according to claim 17, wherein the shaping device introduces the fundamental tone corresponding to the desired pitch which is identical to a pitch of the new sinusoidal components.

19 The apparatus according to claim 17, wherein the shaping device comprises a comb filter having a series of peaks of pass frequencies corresponding to a series of the fundamental

tone and the overtones for filtering the new residual components along a frequency axis.

20 The apparatus according to claim 17, wherein the shaping device comprises a comb filter having a delay loop creating a time delay equivalent to an inverse of the desired pitch for filtering the residual components along a time axis so as to introduce the fundamental tone and the overtones.

21. An apparatus for converting an input voice signal into an output voice signal by modifying a spectral shape, the apparatus comprising:

an input device that provides the input voice signal containing wave components;

an separating device that separates sinusoidal ones of the wave components from the input voice signal such that each sinusoidal wave component is identified by a pair of a frequency and an amplitude;

a computing device that computes a spectral shape of the input voice signal based on a set of the separated sinusoidal wave components such that the spectral shape represents an envelope having a series of break points corresponding to the pairs of the frequencies and the amplitudes of the sinusoidal wave components;

a modifying device that modifies the spectral shape to form a new spectral shape having a modified envelope;

a generating device that selects a series of points along the modified envelope of the new spectral shape and that generates a set of new sinusoidal wave components each identified by each pair of a frequency and an amplitude, which corresponds to each of the series of the selected points; and

an output device that produces the output voice signal based on the set of the new sinusoidal wave components.

22. The apparatus according to claim 21, wherein the output device produces the output voice signal based on the set of the new sinusoidal wave components and residual wave components, which are a part of the wave components of the input voice signal other than the sinusoidal wave components.

23. The apparatus according to claim 21, wherein the modifying device forms the new spectral shape by shifting the envelope along an axis of the frequency on a coordinates system of the frequency and the amplitude.

24. The apparatus according to claim 21, wherein the modifying device forms the new spectral shape by changing a slope of the envelope.

25. The apparatus according to claim 21, wherein the generating device comprises a first section that determines a series of frequencies according to a specific pitch of the

output voice signal, and a second section that selects the series of the points along the modified envelope in terms of the series of the determined frequencies, thereby generating the set of the new sinusoidal wave components corresponding to the series of the selected points and having the determined frequencies.

26. The apparatus according to claim 21, wherein the modifying device modifies the spectral shape to form the new spectral shape according to a specific pitch of the output voice signal such that a modification degree of the frequency or the amplitude of the spectral shape is determined in function of the specific pitch of the output voice signal.

27. The apparatus according to claim 26, further comprising a vibrating device that periodically varies the specific pitch of the output voice signal.

28. The apparatus according to claim 21, wherein the output device produces a plurality of the output voice signals having different pitches, and wherein the modifying device modifies the spectral shape to form a plurality of the new spectral shapes in correspondence with the different pitches of the plurality of the output voice signals.

29. The apparatus according to claim 21, wherein the generating device comprises a first section that selects the

series of the points along the modified envelope of the new spectral shape in which each selected point is denoted by a pair of a frequency and an normalized amplitude calculated using a mean amplitude of the sinusoidal wave components of the input voice signal, and a second section that generates the set of the new sinusoidal wave components in correspondence with the series of the selected points such that each new sinusoidal wave component has a frequency and an amplitude calculated from the corresponding normalized amplitude with using a specific mean amplitude of the new sinusoidal wave components of the output voice signal.

30. The apparatus according to claim 29, further comprising a vibrating device that periodically varies the specific mean amplitude of the new sinusoidal wave components of the output voice signal.

31. An apparatus for converting an input voice signal into an output voice signal dependently on a predetermined pitch of the output voice signal, the apparatus comprising:

an input device that provides the input voice signal containing wave components;

an separating device that separates sinusoidal ones of the wave components from the input voice signal such that each sinusoidal wave component is identified by a pair of a frequency and an amplitude;

a computing device that computes a modification amount of at least one of the frequency and the amplitude of the separated sinusoidal wave components according to the predetermined pitch of the output voice signal;

a modifying device that modifies at least one of the frequency and the amplitude of the separated sinusoidal wave components by the computed modification amount to thereby form new sinusoidal wave components; and

an output device that produces the output voice signal based on the new sinusoidal wave components.

32. An apparatus for discriminating between a voiced state and an unvoiced state at each frame of a voice signal having a waveform oscillating around a zero level with a variable energy, the apparatus comprising:

a zero-cross detecting device that detects a zero-cross point at which the waveform of the voice signal crosses the zero level and that counts a number of the zero-cross points detected within each frame;

an energy detecting device that detects the energy of the voice signal per each frame; and

an analyzing device operative at each frame to determine that the voice signal is placed in the unvoiced state, when the counted number of the zero-cross points is equal to or greater than a lower zero-cross threshold and is smaller than an upper zero-cross threshold, and when the detected energy of the voice signal is equal to or greater than a lower

energy threshold and is smaller than an upper energy threshold.

33. The apparatus according to claim 32, wherein the analyzing device determines that the voice signal is placed in the unvoiced state when the counted number of the zero-cross points is equal to or greater than the upper zero-cross threshold regardless of the detected energy, and determines that the voice signal is placed in a silent state other than the voiced state and the unvoiced state when the detected energy of the voice signal is smaller than the lower energy threshold regardless of the counted number of the zero-cross points.

34. The apparatus according to claim 32, wherein the zero-cross detecting device counts the number of the zero-cross points in terms of a zero-cross factor calculated by dividing the number of the zero-crossing points by a number of sample points of the voice signal contained in one frame, and wherein the energy detecting device detects the energy in terms of an energy factor calculated by accumulating absolute energy values at the sample points throughout one frame and further by dividing the accumulated results by the number of the sample points of the voice signal contained in one frame the.

35. An apparatus for discriminating between a voiced state and an unvoiced state at each frame of a voice signal, the apparatus comprising:

a wave detecting device that processes each frame of the voice signal to detect therefrom a plurality of sinusoidal wave components, each of which is identified by a pair of a frequency and an amplitude;

a separating device that separates the detected sinusoidal wave components into a higher frequency group and a lower frequency group at each frame by comparing the frequency of each sinusoidal wave component with a predetermined reference frequency; and

an analyzing device operative at each frame to determine whether the voice signal is placed in the voiced state or the unvoiced state based on an amplitude related to at least one sinusoidal wave component belonging to the higher frequency group.

36. The apparatus according to claim 35, wherein the analyzing device determines that the voice signal is placed in the unvoiced state when a sinusoidal wave component having the greatest amplitude belongs to the higher frequency group.

37. The apparatus according to claim 35, wherein the analyzing device determines whether the voice signal is placed in the voiced state or the unvoiced state based on a ratio of a mean amplitude of the sinusoidal wave components

belonging to the higher frequency group relative to a mean amplitude of the sinusoidal wave components belonging to the lower frequency group.

38. An apparatus for discriminating between a voiced state and an unvoiced state at each frame of a voice signal having a waveform composed of sinusoidal wave components and oscillating around a zero level with a variable energy, the apparatus comprising:

a zero-cross detecting device that detects a zero-cross point at which the waveform of the voice signal crosses the zero level and that counts a number of the zero-cross points detected within each frame;

an energy detecting device that detects the energy of the voice signal per each frame;

a first analyzing device operative at each frame to determine that the voice signal is placed in the unvoiced state, when the counted number of the zero-cross points is equal to or greater than a lower zero-cross threshold and is smaller than an upper zero-cross threshold, and when the detected energy of the voice signal is equal to or greater than a lower energy threshold and is smaller than an upper energy threshold;

a wave detecting device that processes each frame of the voice signal to detect therefrom a plurality of sinusoidal wave components, each of which is identified by a pair of a frequency and an amplitude;

a separating device that separates the detected sinusoidal wave components into a higher frequency group and a lower frequency group at each frame by comparing the frequency of each sinusoidal wave component with a predetermined reference frequency; and

a second analyzing device operative at each frame when the first analyzing device does not determine that the voice signal is placed in the unvoiced state for determining whether the voice signal is placed in the voiced state or the unvoiced state based on an amplitude related to at least one sinusoidal wave component belonging to the higher frequency group.

39. The apparatus according to claim 38, wherein the first analyzing device determines that the voice signal is placed in the unvoiced state when the counted number of the zero-cross points is equal to or greater than the upper zero-cross threshold regardless of the detected energy, and determines that the voice signal is placed in a silent state other than the voiced state and the unvoiced state when the detected energy of the voice signal is smaller than the lower energy threshold regardless of the counted number of the zero-cross points.

40. A method of converting an input voice signal into an output voice signal according to a target voice signal, the method comprising the steps of:

providing the input voice signal composed of an original sinusoidal component and an original residual component other than the original sinusoidal component;

extracting original attribute data from at least the sinusoidal component of the input voice signal, the original attribute data being characteristic of the input voice signal;

synthesizing new attribute data based on both of the original attribute data derived from the input voice signal and target attribute data being characteristic of the target voice signal composed of a target sinusoidal component and a target residual component other than the sinusoidal component, the target attribute data being derived from at least the target sinusoidal component; and

producing the output voice signal based on the new attribute data and either of the original residual component and the target residual component.

41. A method of converting an input voice signal into an output voice signal according to a target voice signal, the method comprising the steps of:

providing the input voice signal composed of original sinusoidal components and original residual components other than the original sinusoidal components;

separating the original sinusoidal components and the original residual components from each other;

modifying the original sinusoidal components based on target sinusoidal components contained in the target voice signal so as to form new sinusoidal components having a first pitch;

modifying the original residual components based on target residual components contained in the target voice signal other than the target sinusoidal components so as to form new residual components having a second pitch;

shaping the new residual components by removing therefrom a fundamental tone corresponding to the second pitch and overtones of the fundamental tone; and

combining the new sinusoidal components and the shaped new residual components with each other so as to produce the output voice signal having the first pitch.

42 The method according to claim 41, wherein the step of shaping comprises removing the fundamental tone corresponding to the second pitch which is identical to one of a pitch of the original sinusoidal components, a pitch of the target sinusoidal components, and a pitch of the new sinusoidal components.

43. A method of converting an input voice signal into an output voice signal according to a target voice signal, the method comprising the steps of:

providing the input voice signal composed of original sinusoidal components and original residual components other than the original sinusoidal components;

separating the original sinusoidal components and the original residual components from each other;

modifying the original sinusoidal components based on target sinusoidal components contained in the target voice signal so as to form new sinusoidal components;

modifying the original residual components based on target residual components contained in the target voice signal other than the target sinusoidal components so as to form new residual components;

shaping the new residual components by introducing thereinto a fundamental tone and overtones of the fundamental tone corresponding to a desired pitch; and

combining the new sinusoidal components and the shaped new residual components with each other so as to produce the output voice signal.

44 The method according to claim 43, wherein the step of shaping comprises introducing the fundamental tone corresponding to the desired pitch which is identical to a pitch of the new sinusoidal components.

45. A method of converting an input voice signal into an output voice signal by modifying a spectral shape, the method comprising the steps of:

providing the input voice signal containing wave components;

separating sinusoidal ones of the wave components from the input voice signal such that each sinusoidal wave component is identified by a pair of a frequency and an amplitude;

computing a spectral shape of the input voice signal based on a set of the separated sinusoidal wave components such that the spectral shape represents an envelope having a series of break points corresponding to the pairs of the frequencies and the amplitudes of the sinusoidal wave components;

modifying the spectral shape to form a new spectral shape having a modified envelope;

selecting a series of points along the modified envelope of the new spectral shape;

generating a set of new sinusoidal wave components each identified by each pair of a frequency and an amplitude, which corresponds to each of the series of the selected points; and

producing the output voice signal based on the set of the new sinusoidal wave components.

46. The method according to claim 45, wherein the step of producing comprises producing the output voice signal based on the set of the new sinusoidal wave components and residual wave components, which are a part of the wave components of

the input voice signal other than the sinusoidal wave components.

47. A method of converting an input voice signal into an output voice signal dependently on a predetermined pitch of the output voice signal, the method comprising the steps of:

providing the input voice signal containing wave components;

separating sinusoidal ones of the wave components from the input voice signal such that each sinusoidal wave component is identified by a pair of a frequency and an amplitude;

computing a modification amount of at least one of the frequency and the amplitude of the separated sinusoidal wave components according to the predetermined pitch of the output voice signal;

modifying at least one of the frequency and the amplitude of the separated sinusoidal wave components by the computed modification amount to thereby form new sinusoidal wave components; and

producing the output voice signal based on the new sinusoidal wave components.

48. A method of discriminating between a voiced state and an unvoiced state at each frame of a voice signal having a waveform oscillating around a zero level with a variable energy, the method comprising the steps of:

detecting a zero-cross point at which the waveform of the voice signal crosses the zero level so as to count a number of the zero-cross points detected within each frame;

detecting the energy of the voice signal per each frame; and

determining at each frame that the voice signal is placed in the unvoiced state, when the counted number of the zero-cross points is equal to or greater than a lower zero-cross threshold and is smaller than an upper zero-cross threshold, and when the detected energy of the voice signal is equal to or greater than a lower energy threshold and is smaller than an upper energy threshold.

49. A method of discriminating between a voiced state and an unvoiced state at each frame of a voice signal, the method comprising the steps of:

processing each frame of the voice signal to detect therefrom a plurality of sinusoidal wave components, each of which is identified by a pair of a frequency and an amplitude;

separating the detected sinusoidal wave components into a higher frequency group and a lower frequency group at each frame by comparing the frequency of each sinusoidal wave component with a predetermined reference frequency; and

determining at each frame whether the voice signal is placed in the voiced state or the unvoiced state based on an

amplitude related to at least one sinusoidal wave component belonging to the higher frequency group.

50. A machine readable medium used in a computer machine having a CPU, the medium containing program instructions executable by the CPU to cause the computer machine for performing a process of converting an input voice signal into an output voice signal according to a target voice signal, the process comprising the steps of:

providing the input voice signal composed of an original sinusoidal component and an original residual component other than the original sinusoidal component;

extracting original attribute data from at least the sinusoidal component of the input voice signal, the original attribute data being characteristic of the input voice signal;

synthesizing new attribute data based on both of the original attribute data derived from the input voice signal and target attribute data being characteristic of the target voice signal composed of a target sinusoidal component and a target residual component other than the sinusoidal component, the target attribute data being derived from at least the target sinusoidal component; and

producing the output voice signal based on the new attribute data and either of the original residual component and the target residual component.

51. A machine readable medium used in a computer machine having a CPU, the medium containing program instructions executable by the CPU to cause the computer machine for performing a process of converting an input voice signal into an output voice signal according to a target voice signal, the process comprising the steps of:

providing the input voice signal composed of original sinusoidal components and original residual components other than the original sinusoidal components;

separating the original sinusoidal components and the original residual components from each other;

modifying the original sinusoidal components based on target sinusoidal components contained in the target voice signal so as to form new sinusoidal components having a first pitch;

modifying the original residual components based on target residual components contained in the target voice signal other than the target sinusoidal components so as to form new residual components having a second pitch;

shaping the new residual components by removing therefrom a fundamental tone corresponding to the second pitch and overtones of the fundamental tone; and

combining the new sinusoidal components and the shaped new residual components with each other so as to produce the output voice signal having the first pitch.

52 The machine readable medium according to claim 51, wherein the step of shaping comprises removing the fundamental tone corresponding to the second pitch which is identical to one of a pitch of the original sinusoidal components, a pitch of the target sinusoidal components, and a pitch of the new sinusoidal components.

53. A machine readable medium used in a computer machine having a CPU, the medium containing program instructions executable by the CPU to cause the computer machine for performing a process of converting an input voice signal into an output voice signal according to a target voice signal, the process comprising the steps of:

providing the input voice signal composed of original sinusoidal components and original residual components other than the original sinusoidal components;

separating the original sinusoidal components and the original residual components from each other;

modifying the original sinusoidal components based on target sinusoidal components contained in the target voice signal so as to form new sinusoidal components;

modifying the original residual components based on target residual components contained in the target voice signal other than the target sinusoidal components so as to form new residual components;

shaping the new residual components by introducing thereinto a fundamental tone and overtones of the fundamental tone corresponding to a desired pitch; and

combining the new sinusoidal components and the shaped new residual components with each other so as to produce the output voice signal.

54 The machine readable medium according to claim 53, wherein the step of shaping comprises introducing the fundamental tone corresponding to the desired pitch which is identical to a pitch of the new sinusoidal components.

55. A machine readable medium used in a computer machine having a CPU, the medium containing program instructions executable by the CPU to cause the computer machine for performing a process of converting an input voice signal into an output voice signal by modifying a spectral shape, the process comprising the steps of:

providing the input voice signal containing wave components;

separating sinusoidal ones of the wave components from the input voice signal such that each sinusoidal wave component is identified by a pair of a frequency and an amplitude;

computing a spectral shape of the input voice signal based on a set of the separated sinusoidal wave components such that the spectral shape represents an envelope having a

series of break points corresponding to the pairs of the frequencies and the amplitudes of the sinusoidal wave components;

modifying the spectral shape to form a new spectral shape having a modified envelope;

selecting a series of points along the modified envelope of the new spectral shape;

generating a set of new sinusoidal wave components each identified by each pair of a frequency and an amplitude, which corresponds to each of the series of the selected points; and

producing the output voice signal based on the set of the new sinusoidal wave components.

56. The machine readable medium according to claim 55, wherein the step of producing comprises producing the output voice signal based on the set of the new sinusoidal wave components and residual wave components, which are a part of the wave components of the input voice signal other than the sinusoidal wave components.

57. A machine readable medium used in a computer machine having a CPU, the medium containing program instructions executable by the CPU to cause the computer machine for performing a process of converting an input voice signal into an output voice signal dependently on a predetermined pitch

of the output voice signal, the process comprising the steps of:

providing the input voice signal containing wave components;

separating sinusoidal ones of the wave components from the input voice signal such that each sinusoidal wave component is identified by a pair of a frequency and an amplitude;

computing a modification amount of at least one of the frequency and the amplitude of the separated sinusoidal wave components according to the predetermined pitch of the output voice signal;

modifying at least one of the frequency and the amplitude of the separated sinusoidal wave components by the computed modification amount to thereby form new sinusoidal wave components; and

producing the output voice signal based on the new sinusoidal wave components.

58. A machine readable medium used in a computer machine having a CPU, the medium containing program instructions executable by the CPU to cause the computer machine for performing a process of discriminating between a voiced state and an unvoiced state at each frame of a voice signal having a waveform oscillating around a zero level with a variable energy, the process comprising the steps of:

detecting a zero-cross point at which the waveform of the voice signal crosses the zero level so as to count a number of the zero-cross points detected within each frame;

detecting the energy of the voice signal per each frame; and

determining at each frame that the voice signal is placed in the unvoiced state, when the counted number of the zero-cross points is equal to or greater than a lower zero-cross threshold and is smaller than an upper zero-cross threshold, and when the detected energy of the voice signal is equal to or greater than a lower energy threshold and is smaller than an upper energy threshold.

59. A machine readable medium used in a computer machine having a CPU, the medium containing program instructions executable by the CPU to cause the computer machine for performing a process of discriminating between a voiced state and an unvoiced state at each frame of a voice signal, the process comprising the steps of:

processing each frame of the voice signal to detect therefrom a plurality of sinusoidal wave components, each of which is identified by a pair of a frequency and an amplitude;

separating the detected sinusoidal wave components into a higher frequency group and a lower frequency group at each frame by comparing the frequency of each sinusoidal wave component with a predetermined reference frequency; and

determining at each frame whether the voice signal is placed in the voiced state or the unvoiced state based on an amplitude related to at least one sinusoidal wave component belonging to the higher frequency group.